

Campuses & the OSG Computing Landscape

Rob Gardner
Chander Sehgal
Brian Bockelman

Substituting for
David Swanson

HTC
DHTC

The OSG fabric of services

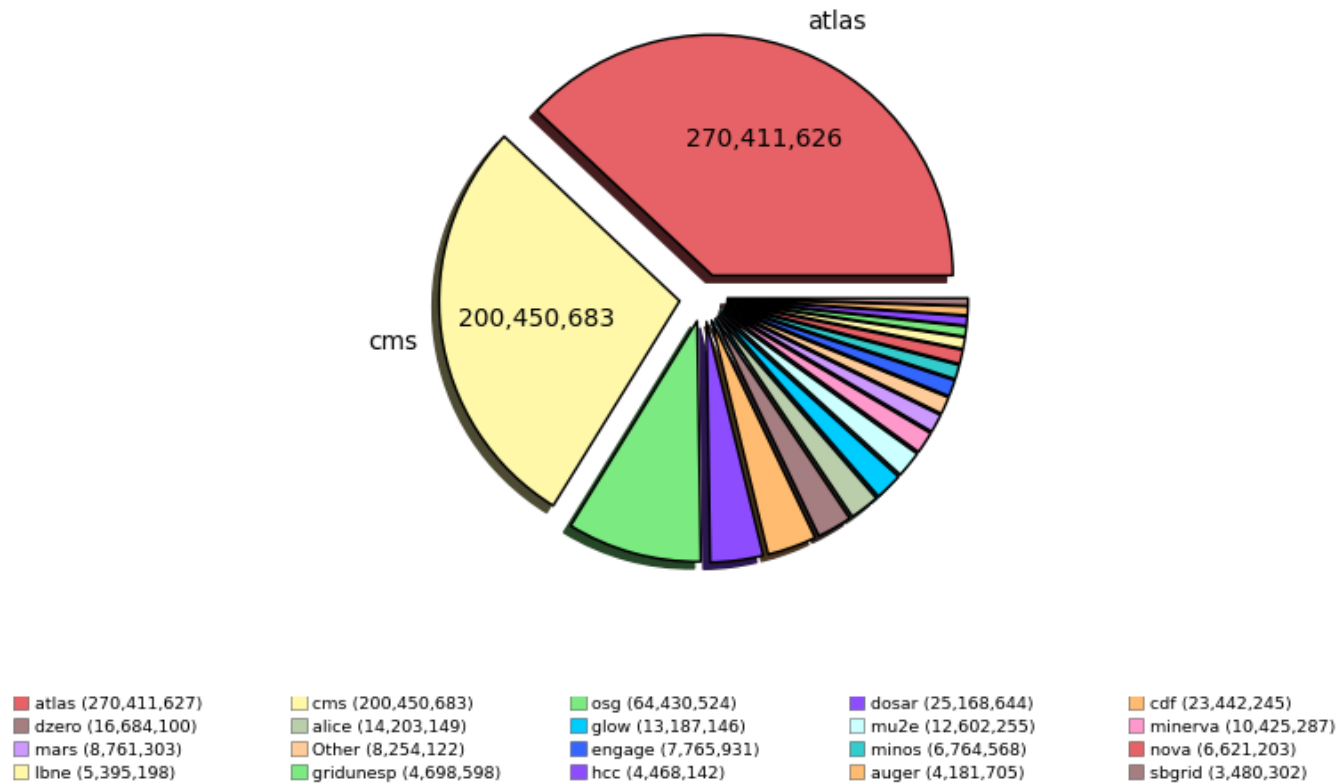
- The leading **distributed** high throughput computing service in the US
 - 104k cores, 75.6 PB, 123 compute endpoints
- Well connected
 - Most sites on OSG have 10 Gbps or greater to I2 or ESnet, many upgrading, plus SciDMZs
 - (at least) 2M transfers/day, 1 PB/day
- Friction free
 - The OSG VO with **GlideinWMS** offers transparent access to these resources for small groups

OSG Ecosystem

All OSG Usage for 12 months ending 31-March-2014

Wall Hours by VO (Sum: 711,396,733 Hours)

53 Weeks from Week 13 of 2013 to Week 13 of 2014

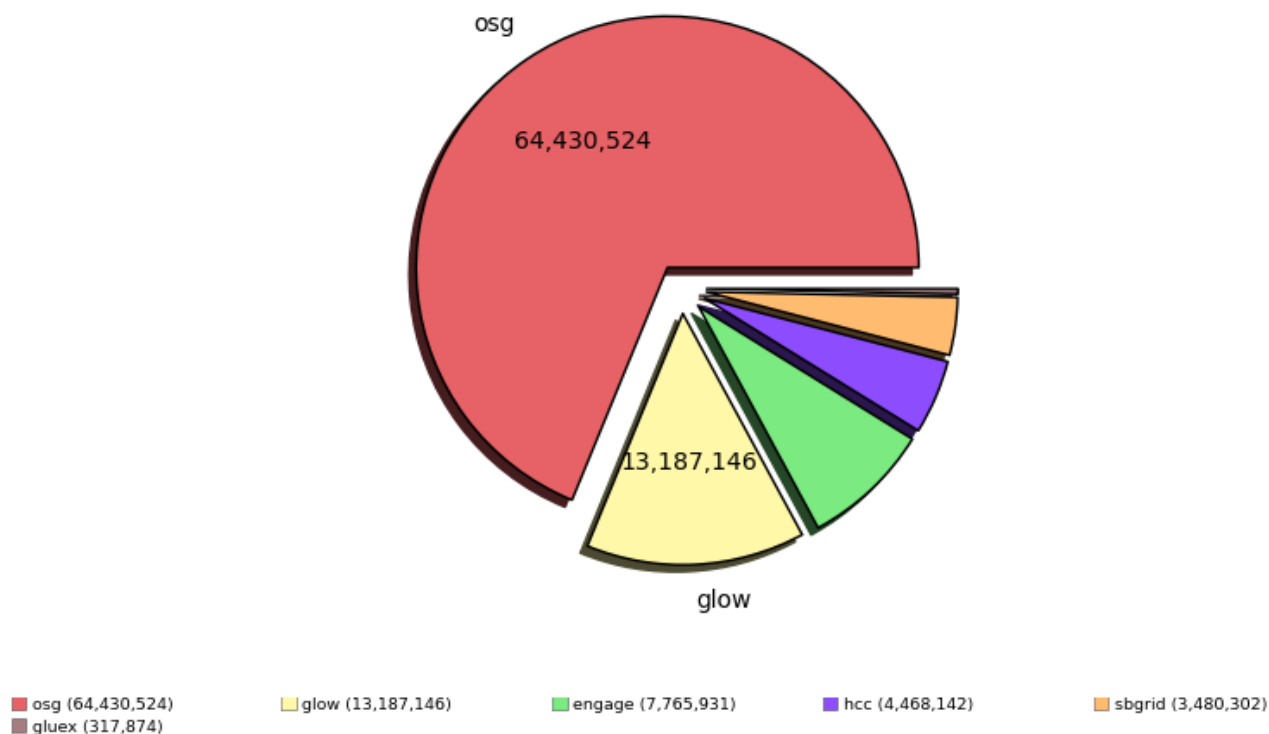


Some of these VOs access opportunistic cycles
e.g. osg, glow, engage, hcc, sbgrid

Opportunistic Ecosystem

Usage by “opportunistic VOs” for 12 months ending 31-March-2014

Wall Hours by VO (Sum: 93,649,919 Hours)
53 Weeks from Week 13 of 2013 to Week 13 of 2014



Of these, the OSG VO provides access to US researchers who are not already affiliated with an existing community in OSG

OSG: the Opportunity VO

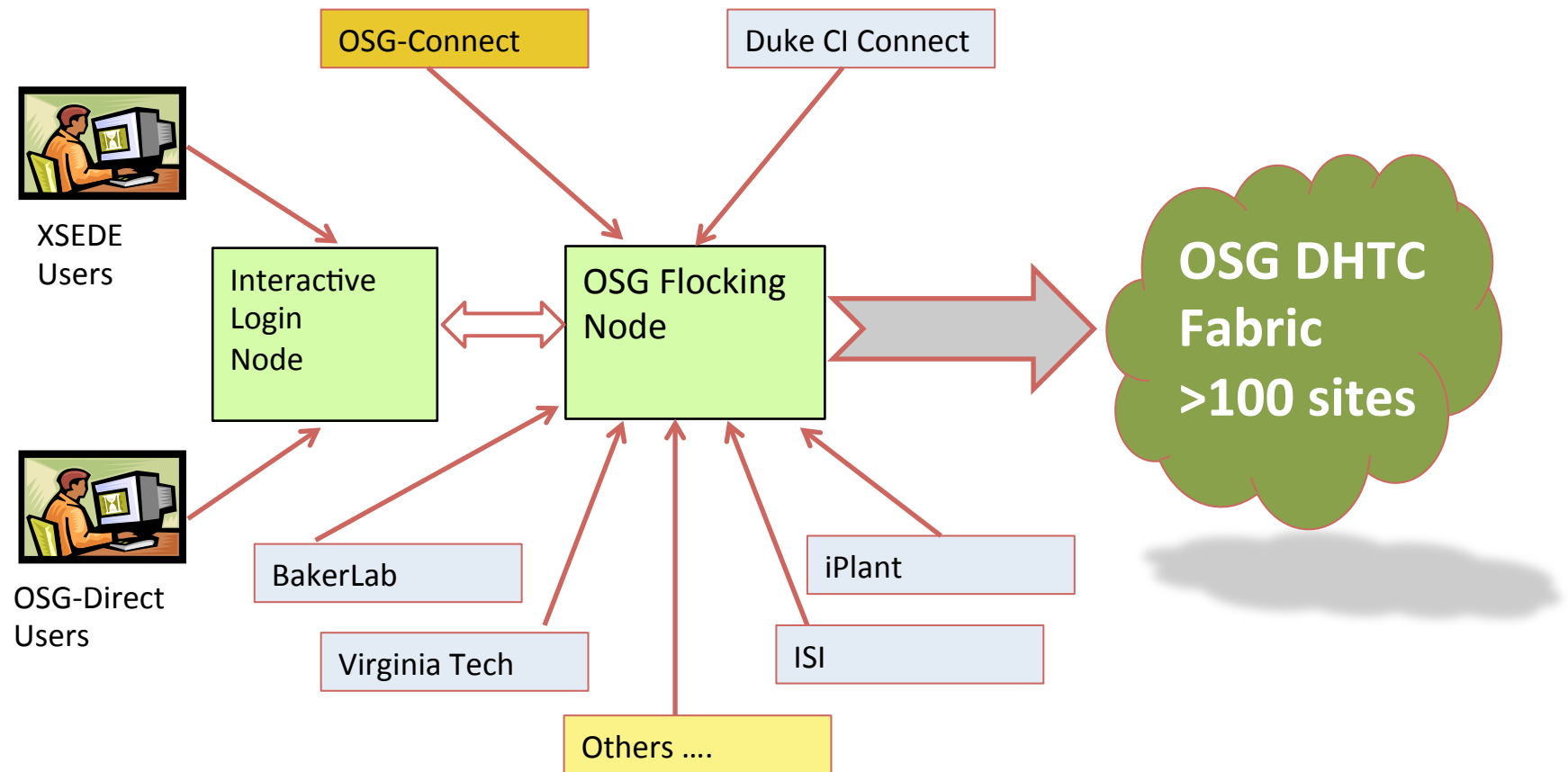
The **OSG VO does not own any computing resources** and only exists to harvest unused cycles at OSG sites (Opportunistic cycles) and make them available to researchers who are not already affiliated with an OSG VO.

For the 12 months ending 31-March-2014, the OSG VO harvested 64.4M hours (from sites by using gWMS) and delivered 57.7M hours to various submit hosts to enable the computing of researchers

Submit Host	Wall Hours
OSG-XD (XSEDE and OSG Direct)**	54,694,294
UCSDgrid	1,104,882
Bakerlab	1,012,264
OSGCONNECT **	870,640
ISI	3,539
LSU	63
Total	57,685,682

** Core OSG Services

Access to OSG via OSG VO



OSG-Direct users

April 2013 to March 2014

Project Name	PI	Institution	Field of Science	Wall Hours
Snowmass	Meenakshi Narain	Brown University	High Energy Physics	8,632,986
SPLINTER	Robert Quick	Indiana University	Medicine	4,601,962
Duke-QGP	Steffen A. Bass	Duke University	Nuclear Physics	2,543,933
ECFA	Meenakshi Narain	Brown University	High Energy Physics	1,744,646
UMich	Paul Wolberg	University of Michigan	Microbiology	1,433,598
Pheno	Stefan Hoeche	SLAC	High Energy Physics	1,108,623
RIT	P. Stanislaw Radziszowski	Rochester Institute of Technology	Computer Science	721,291
UPRRP-MR	Steven Massey	Universidad de Puerto Rico (UPRRP)	Bioinformatics	714,359
IU-GALAXY	Robert Quick	Indiana University	Bioinformatics	640,484
DetectorDesign	John Strologas	University of New Mexico	Medical Imaging	451,803
EIC	Tobias Toll	Brookhaven National Laboratory	Accelerator Physics	410,594
OSG-Staff	Chander Sehgal	Fermilab	Computer Science	43,948
DeerDisease	Lene Jung Kjaer	Southern Illinois University	Biological Sciences	28,599
SNOplus	Joshua R Klein	University of Pennsylvania	Physics - Neutrino	489
P0-LBNE	Maxim Potekhin	Brookhaven National Laboratory	Physics - Neutrino	17
BNLPET	Martin Purschke	Brookhaven National Laboratory	Medical Imaging	1
Total		16 users		23,077,333

XSEDE use of OSG

April 2013 to March 2014

Project Name	PI	Institution	Field of Science	Wall Hours
TG-IBN130001	Donald Krieger	University of Pittsburgh	Biological Sciences	29,495,083
TG-PHY120014	Qaisar Shafi	University of Delaware	Physics	528,458
TG-TRA100004	Andrew Ruether	Swarthmore College	Other	444,374
TG-DMR130036	Emanuel Gull	University of Michigan	Materials Research	318,768
TG-MCB100109	Lillian Chong	University of Pittsburgh	Molecular Biosciences	264,362
TG-CHE130091	Paul Siders	University of Minnesota; Duluth	Chemistry	86,280
TG-ATM130015	Phillip Anderson	University of Texas at Dallas	Atmospheric Sciences	77,169
		University of Massachusetts;		
TG-IRI130016	Joseph Cohen	Boston	Information; Robotics; and Intelligent Systems	70,536
TG-DMS120024	Benjamin Ong	Michigan State University	Mathematical Sciences	68,908
		Massachusetts Institute of		
		Technology	Chemistry	58,355
TG-CHE130103	Jeremy Moix	University of Texas at Dallas	Atmospheric Sciences	39,971
TG-ATM130009	Phillip Anderson	University of Texas at Dallas	Atmospheric Sciences	39,971
TG-MCB090163	Michael Hagan	Brandeis University	Molecular Biosciences	38,590
TG-OCE130029	Yvonne Chan	University of Hawaii; Manoa	Ocean Sciences	31,670
TG-TRA120014	Pol Llovet	Montana State University	Cross-Disciplinary Activities	19,472
TG-IBN130008	Jorden Schossau	Michigan State University	Biological Sciences	16,857
TG-MCB120070	Joseph Hargitai	Albert Einstein College of Medicine	Molecular Biosciences	378
TG-TRA120041	Hanning Chen	George Washington University	Computer and Information Science	231
TG-MCB090174	Shantenu Jha	Rutgers University	Molecular Biosciences	58
TG-PHY110015	Pran Nath	Northeastern University	Physics	37
TG-MCB130072	Robert Quick	Indiana University	Molecular Biosciences	16
TG-CCR120041	Luca Clementi	San Diego Supercomputer Center	Computer and Computation Research	12
	Nancy Wilkins-			
TG-STA110014S	Diehr	University of California-San Diego	Other	5
Total		22 users		31,559,590

Campuses & the National Cyber Ecosystem

- In OSG we've excelled in deploying services to pools to reach stakeholder provisioned resources
- But this is tiny fraction of the potential DHTEC ecosystem

They are out there



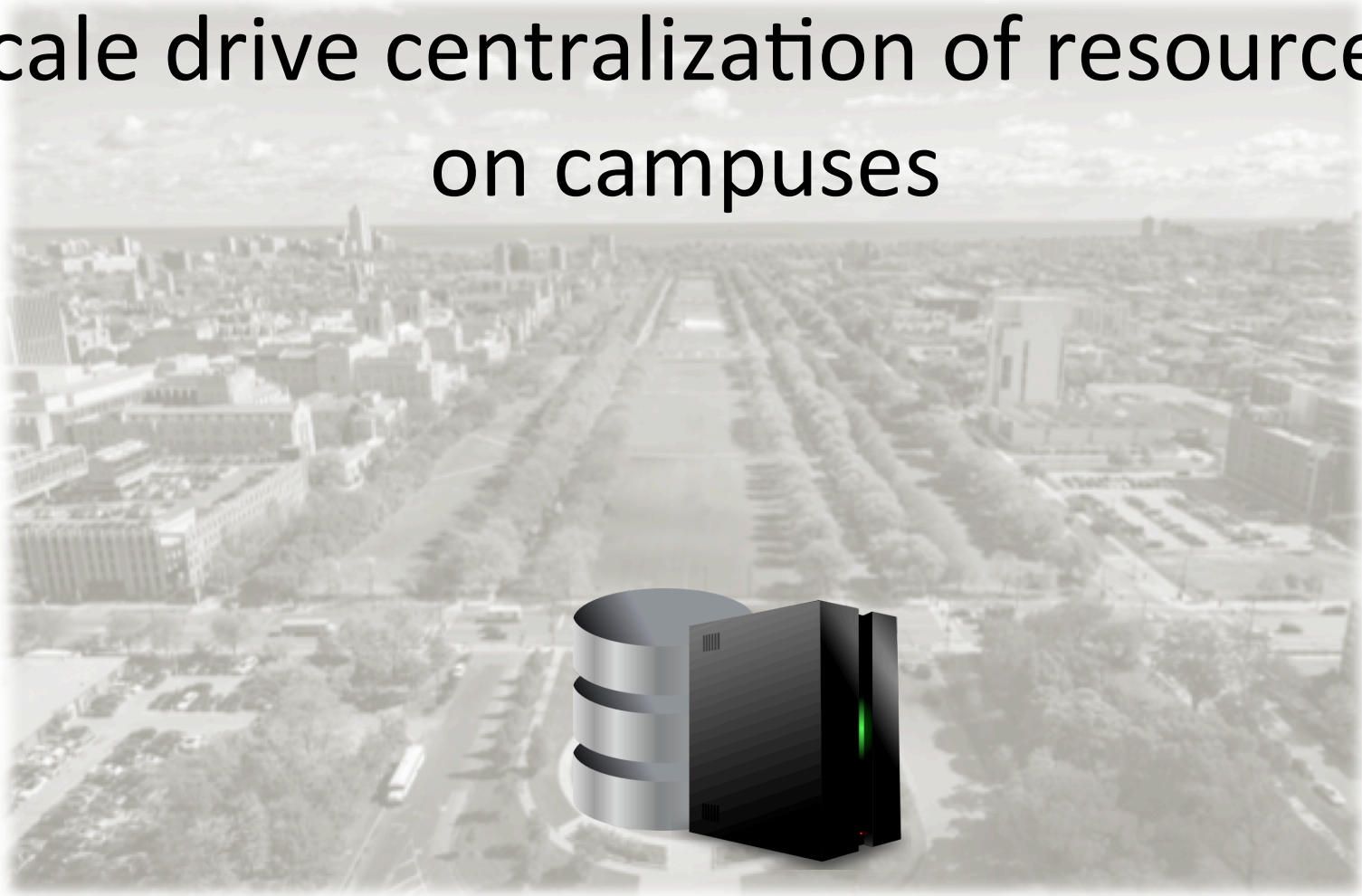
Typical campus: divisions, colleges, departments
➔ distributed PIs and resources



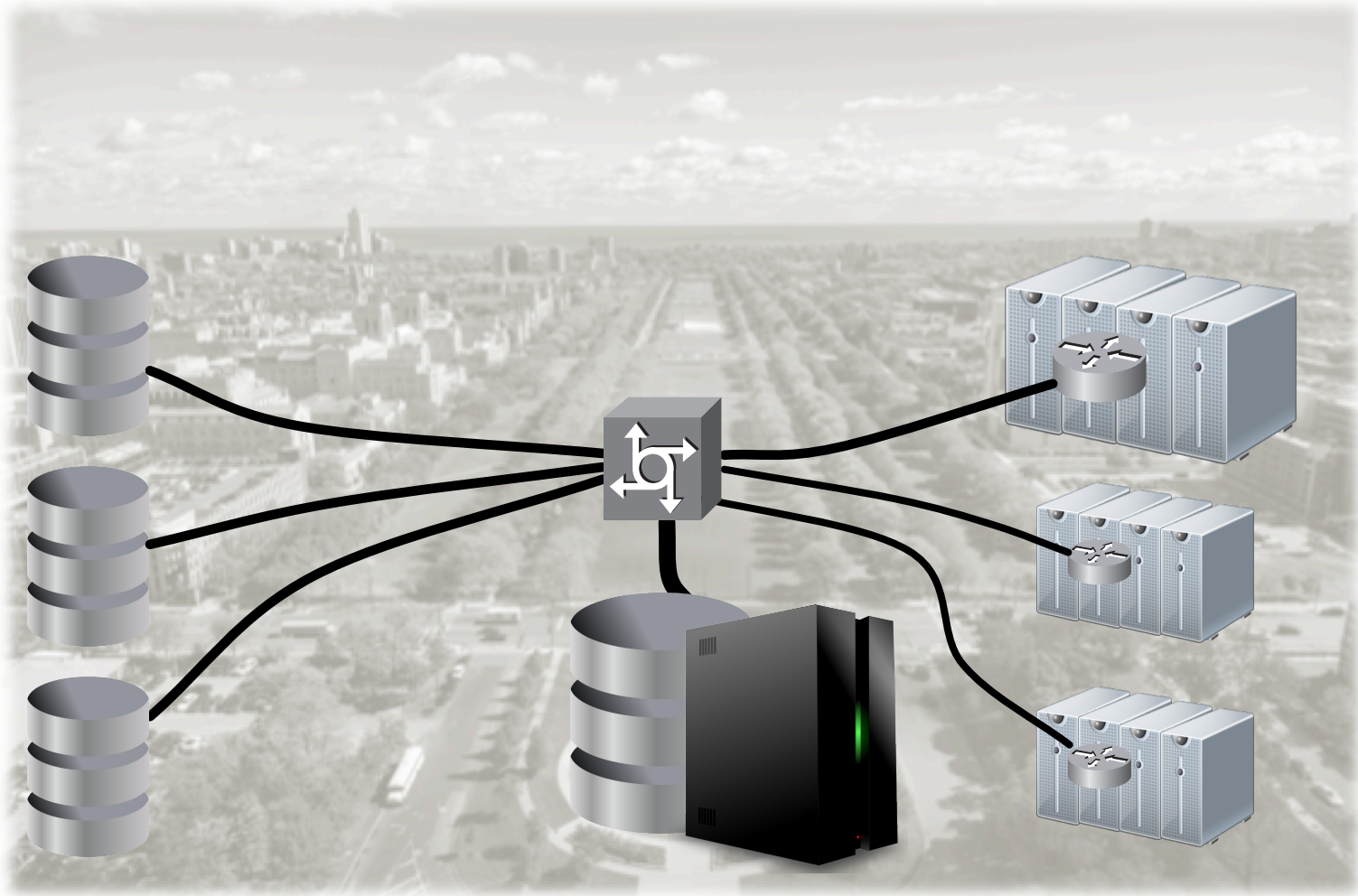
Challenges:
limited budgets, operational costs, increasing
scale of science



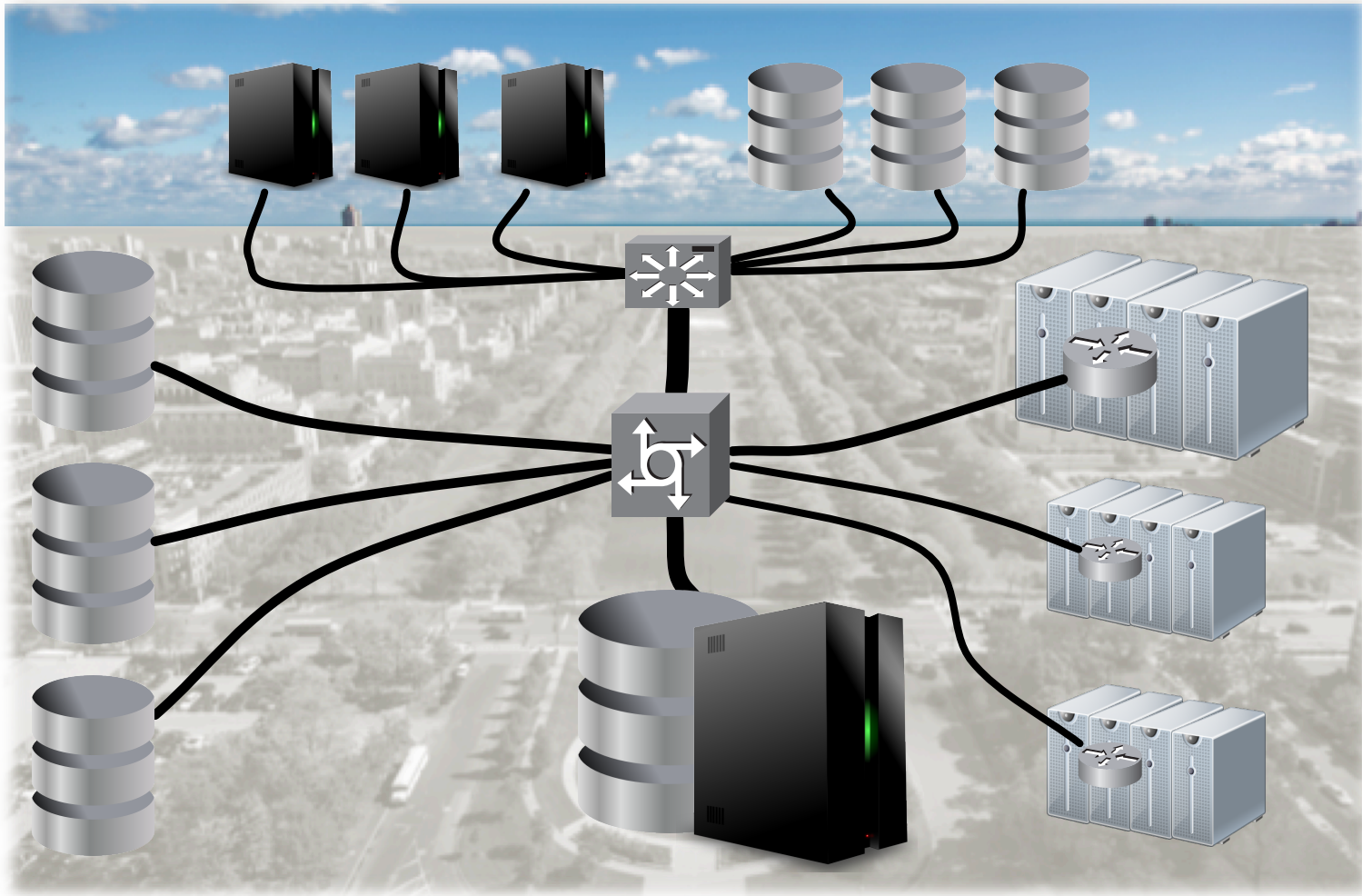
Commodification, cloud technologies,
and practices achieving economies of
scale drive centralization of resources
on campuses



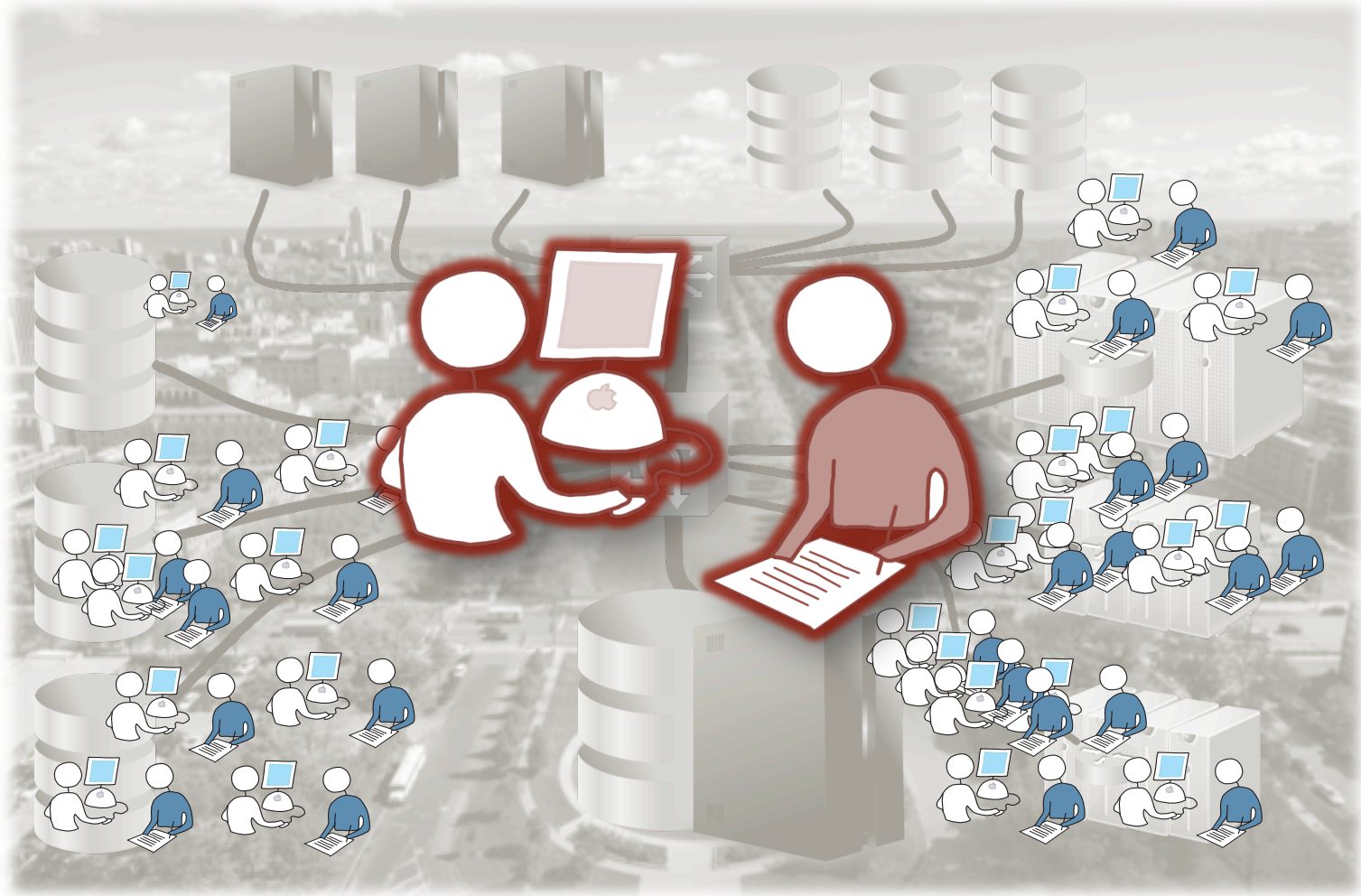
Distributed on Campus



Off campus too (the ecosystem)



Needed: services to make distributed resources
transparent to campus users but **cost effective**
for resource providers



OSG Campus Grids Mission

- Deliver distributed high throughput computing capabilities to campuses while enabling campus researchers with DHTC ready computing tasks
- Create a campus infrastructures community with workshops, online webinars and website to share best practices

Key Approaches

- “Platform of Services” model to facilitate delivery of needed DHTC services for campus researchers
- OSG Connect platform, an interactive job and data service for the OSG

Since Last Year

- CI Connect proposed as a template for creating campus grids and campus bridges
- Two new OSG campus grids built with CI Connect
 - Duke CI Connect (Duke grid, bridged to UC3, OSG)
 - ATLAS Connect (LHC “Tier3” batch analysis pilot)

The OSG Connect Platform

- “Login to the OSG using your campus identity”
- Launched at OSG Campus Infrastructures Community sponsored Workshop at Duke
 - August 26-27, 2013
 - About 25 users signed up using Duke credentials
 - Thousands of tutorial jobs executed on OSG
 - Several Duke OSG Projects created

Components

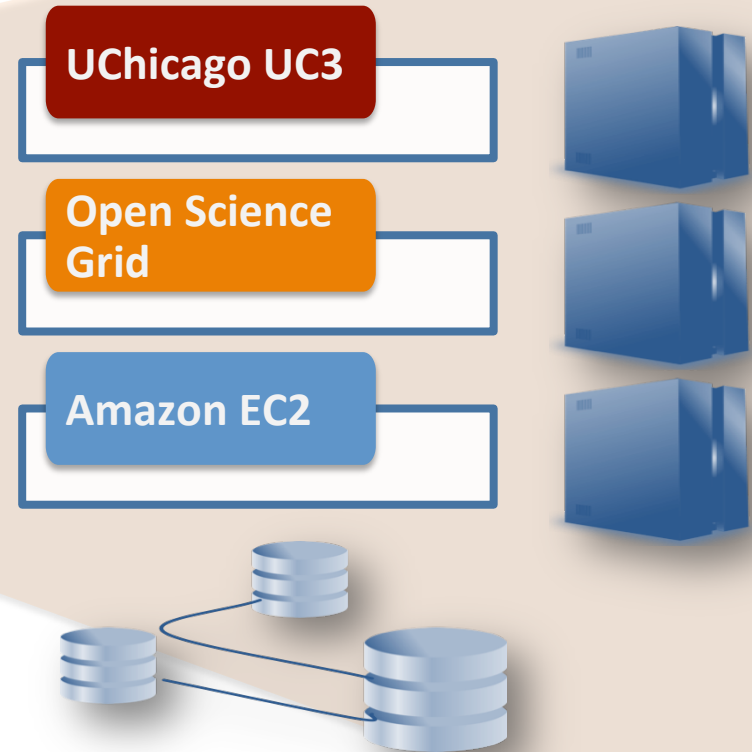
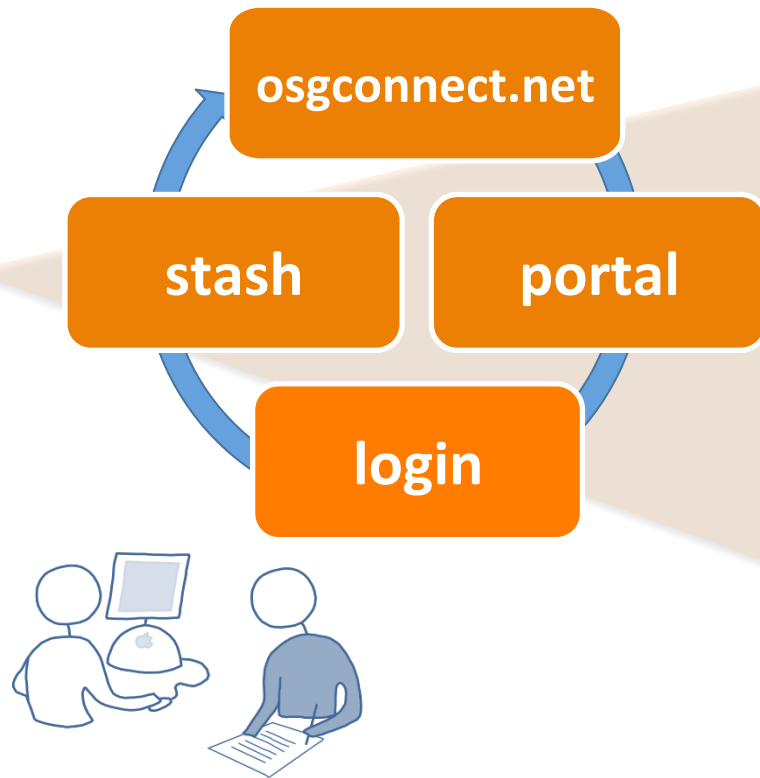
- Leverages Globus, HTCondor, CI-Logon, U-Bolt, Bosco technologies
 - Bundled as instance of a CI Connect service portfolio
 - *Provided as a **Service** to reduce Campus IT load*
- Submit host
 - Flocks to OSG VO front-end, UC3 grid, & Amazon if needed
- Object storage service (90 TB usable)
 - POSIX, Globus Online, http, chirp access protocols
- Accounting (Gratia) and monitoring (Cycle Server) services

Components, cont.

- Inside Science DMZ with 80 Gbps uplink
 - Direct peering I2 & ESnet at CIC OmniPoP (100 Gbps); PerfSonar
 - Co-located with major ATLAS Tier2 center and OSG opportunistic cycle provider
- Full advantage of Globus for reliable file transfer, data sharing
- Integrated web portal & group organization
- User-focused knowledge base (ConnectBook)



osg connect

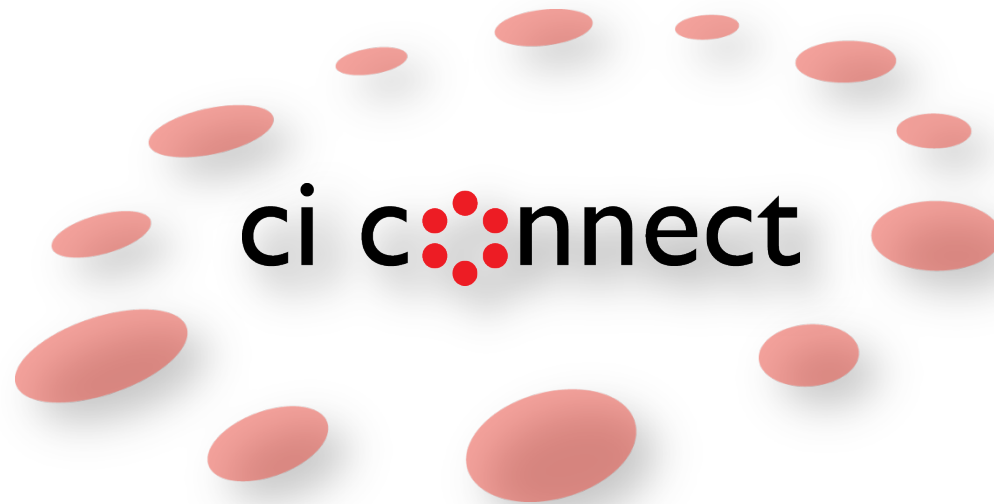


CPU Hours Delivered

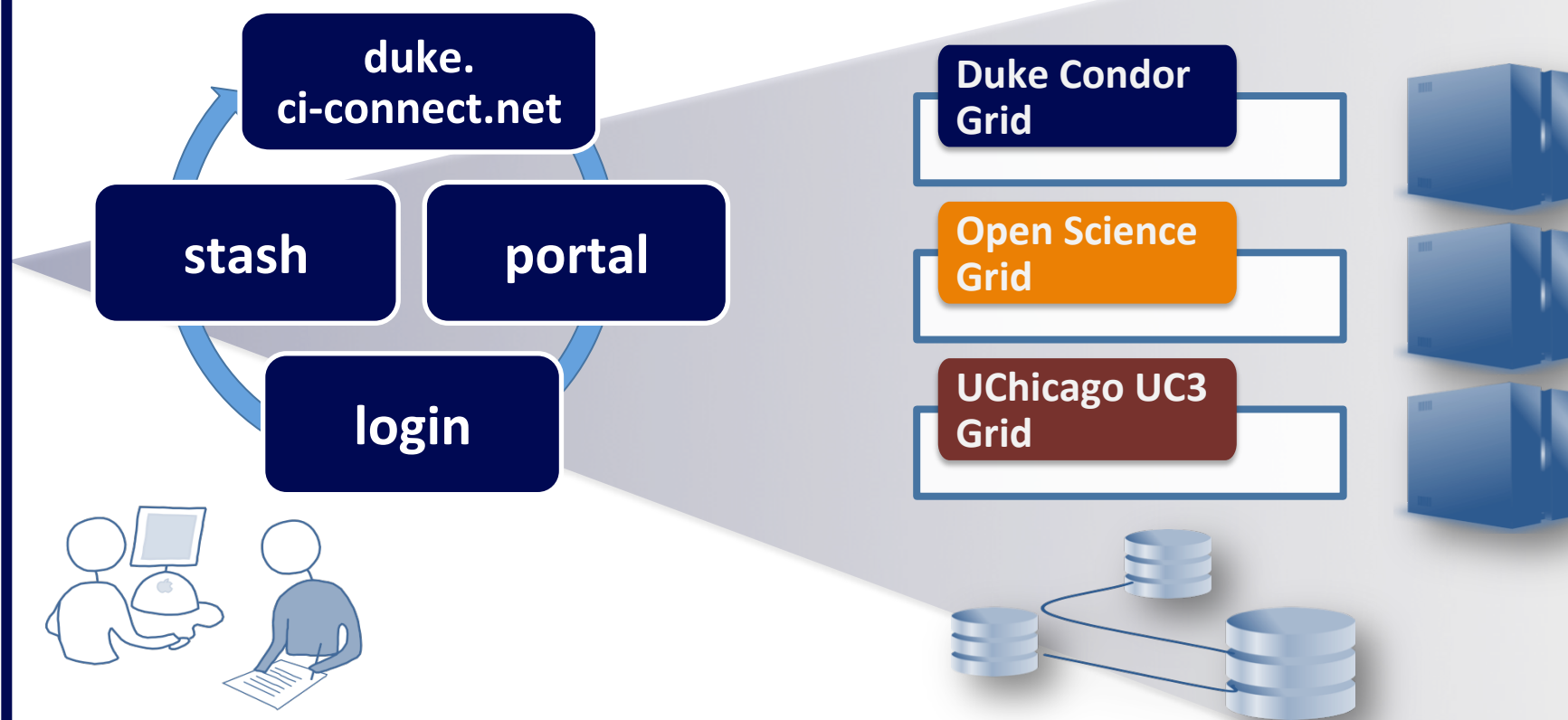
Project Name	Field of Science	Usage Class	Contact	Institution	CPU hours
osg.KnowledgeSys	Psychology	Significant	Michael Culbertson	University of Illinois	383,182
osg.NESCent	Cross disciplinary Evolution Studies	Significant	Fabricia Nascimento	National Evolutionary Synthesis Center	220,547
osg.Staff	N/A	Significant	Rob Gardner	University of Chicago	143,685
osg.PathSpaceHMC	Computational Condensed Matter	Significant	Frank Pinski	University of Cincinnati	60,288
osg.ConnectTrain	Physics	Significant	Various	Various	23,793
osg.Swift	N/A	Significant	Mike Wilde	University of Chicago	11,837
osg.SouthPoleTelescope	Computer Science	Significant	John Carlstrom	University of Chicago	7,940
osg.EvoTheory	Astrophysics	Significant	John Carlstrom	University of Chicago	7,940
osg.EvoTheory	Evolutionary Biology	Some	Christina Burch	University of North Carolina Chapel Hill	264
osg.CompChem	Chemistry	Some	Chaoren Liu	Duke University	229
osg.DBConcepts	English Literature	Some	Richard Jean So	University of Chicago	4
osg.RADICAL	Computer Science	Experimented	Shantenu Jha	University of Chicago	1
osg.GlassySystems	Chemistry	Experimented	David Reichman	Columbia University	1
osg.CompNeuro	Neuroscience	Experimented	Po-He Tseng	Duke University	0
osg.AMFORA			Zhao Zhang (Ian Foster)		
osg.UChicago-RCC	Computer Science	Experimented	Foster)	University of Chicago	0
osg.UChicago-RCC	Research Computing Center	Experimented	Birali Runesha	University of Chicago	0
osg.NRELMatDB	Material Science	Experimented	Steve Sullivan	National Renewable Energy Laboratory	0
osg.BioStat	Bioninformatics	No accounting	Janice McCarthy	Duke University	0
osg.PlantBio	Plant Biology	No accounting	Joy Bergleson	University of Chicago	0
osg.RDCEP	Economics	No accounting	Ian Foster	University of Chicago	0
OSG Connect	All	Significant	Rob Gardner	All	851,772
Duke CI Connect	Campus grid	Significant	Tom Millege	Duke University	1,176,498
ATLAS Connect	Community grid	Significant	Rob Gardner	US ATLAS	255,077

Total usage 2,283,347

OSG Connect → Suggests Campus Grids as a Service



ci-connect.net built on Globus Platform and HTCondor

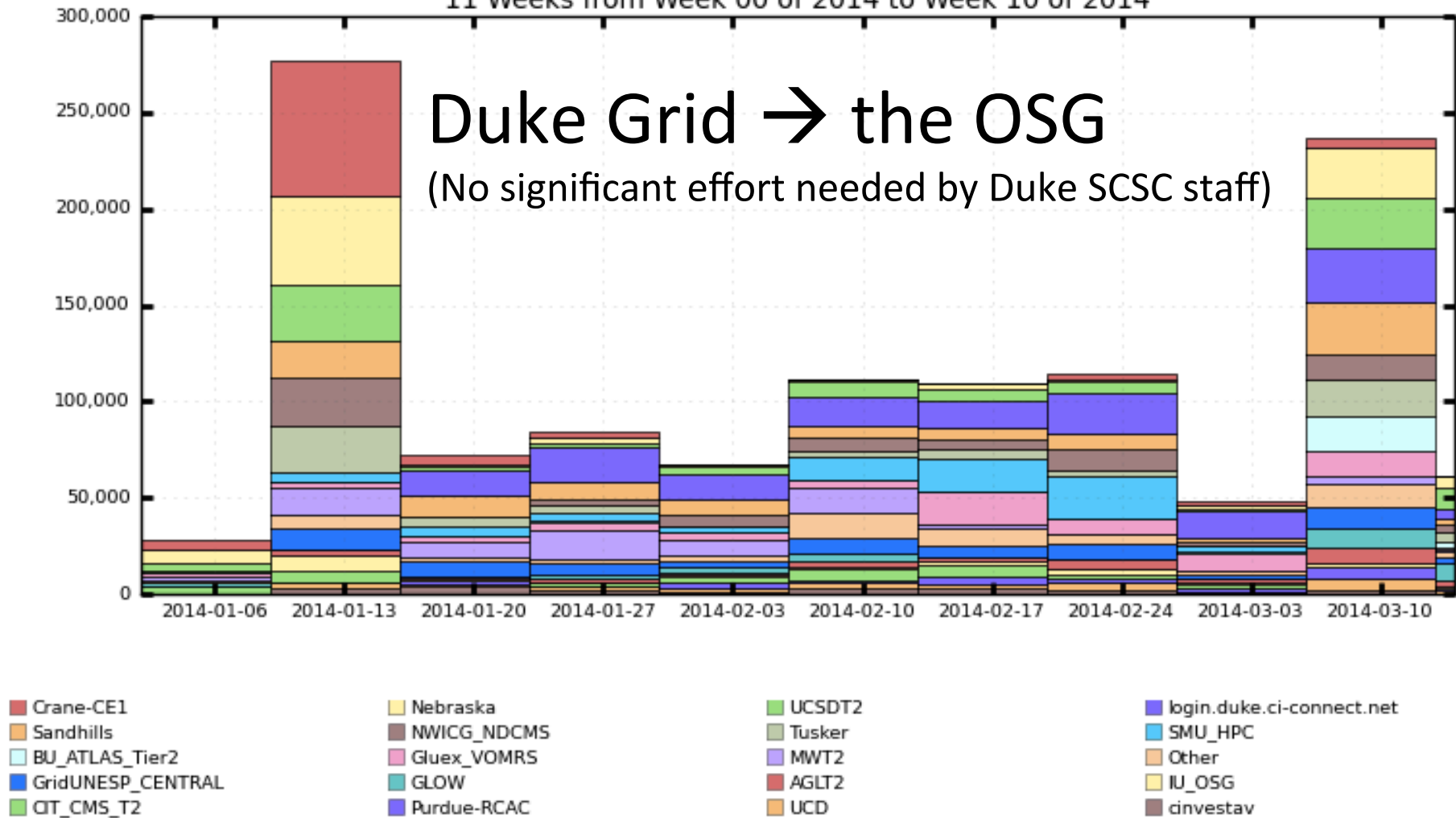


WMS Hours Spent on Jobs By Facility (Glidein)

11 Weeks from Week 00 of 2014 to Week 10 of 2014

Duke Grid → the OSG

(No significant effort needed by Duke SCSC staff)



Maximum: 276,604 , Minimum: 28,237 , Average: 110,106 , Current: 61,593

CI Connect-based Campus Grids

- **ATLAS
Connect**

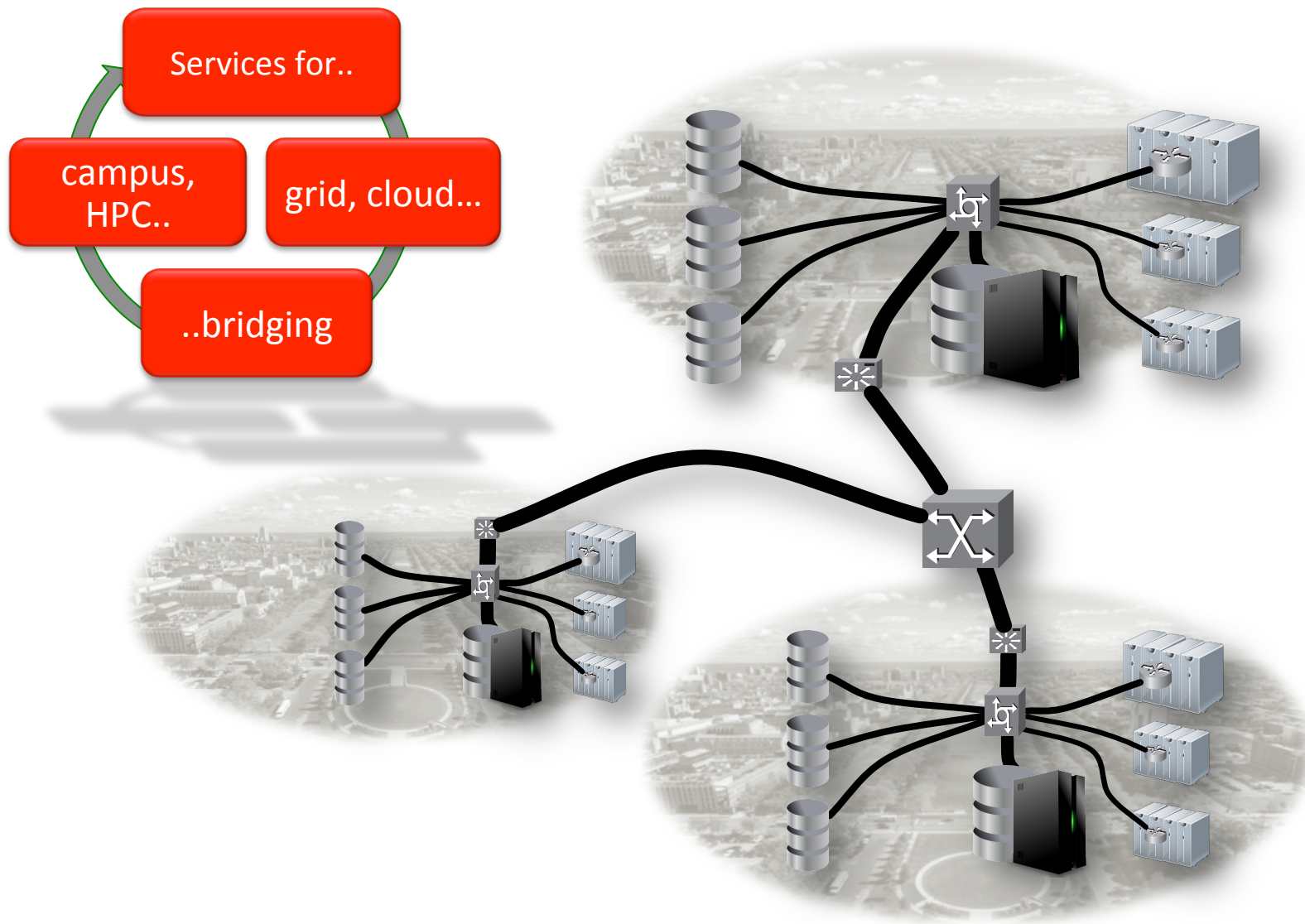
**Potential to
connect 44
US institutions:
users & Tier 3
infrastructure**

Project Name	Usage Class	CPU hours
ATLAS-ORG-UCHICAGO	Significant usage	157,840
ATLAS-WG-HIGGS	Significant usage	30,059
ATLASCONNECT	Significant usage	24,037
ATLAS	Significant usage	20,042
ATLAS-ORG-FRESNO-STATE	Significant usage	7,246
ATLAS-ORG-UTEXAS	Significant usage	6,478
ATLAS-WG-SUSY	Some usage	1,857
CONNECT	Training	1,277
STAFF	Test and benchmark	6,242
Total		255,077

- **Duke
CI Connect**

Project Name	Usage Class	CPU hours
DUKE-4FERMION	Significant usage	747,931
DUKE	Significant usage	422,230
ATLASCONNECT	Significant usage	5,289
DUKE-QGP	Experimented	0
CONNECTTRAIN	Training	932
OSG-STAFF	Test and benchmark	116
Total		1,176,498

CI Bridging for Campuses

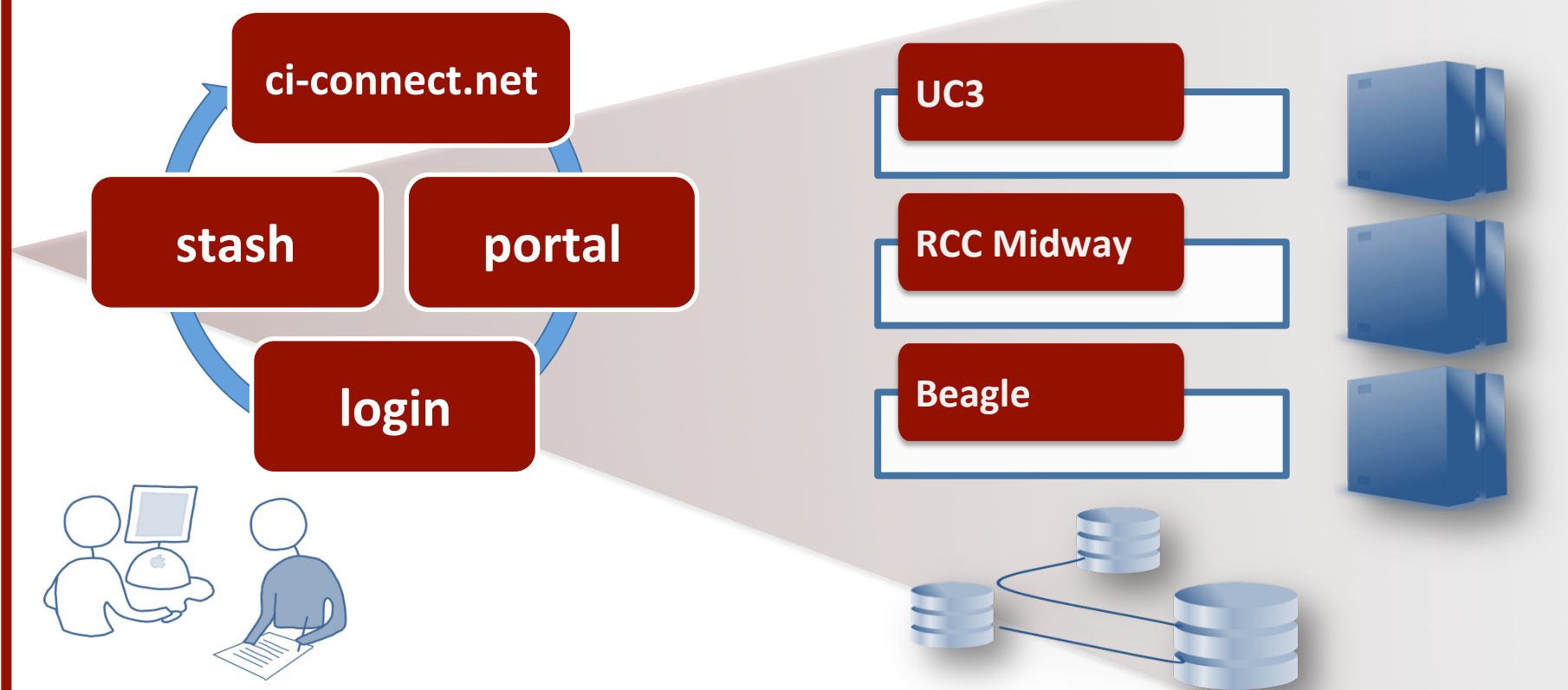




THE UNIVERSITY OF
CHICAGO

CI CONNECT

In preparation

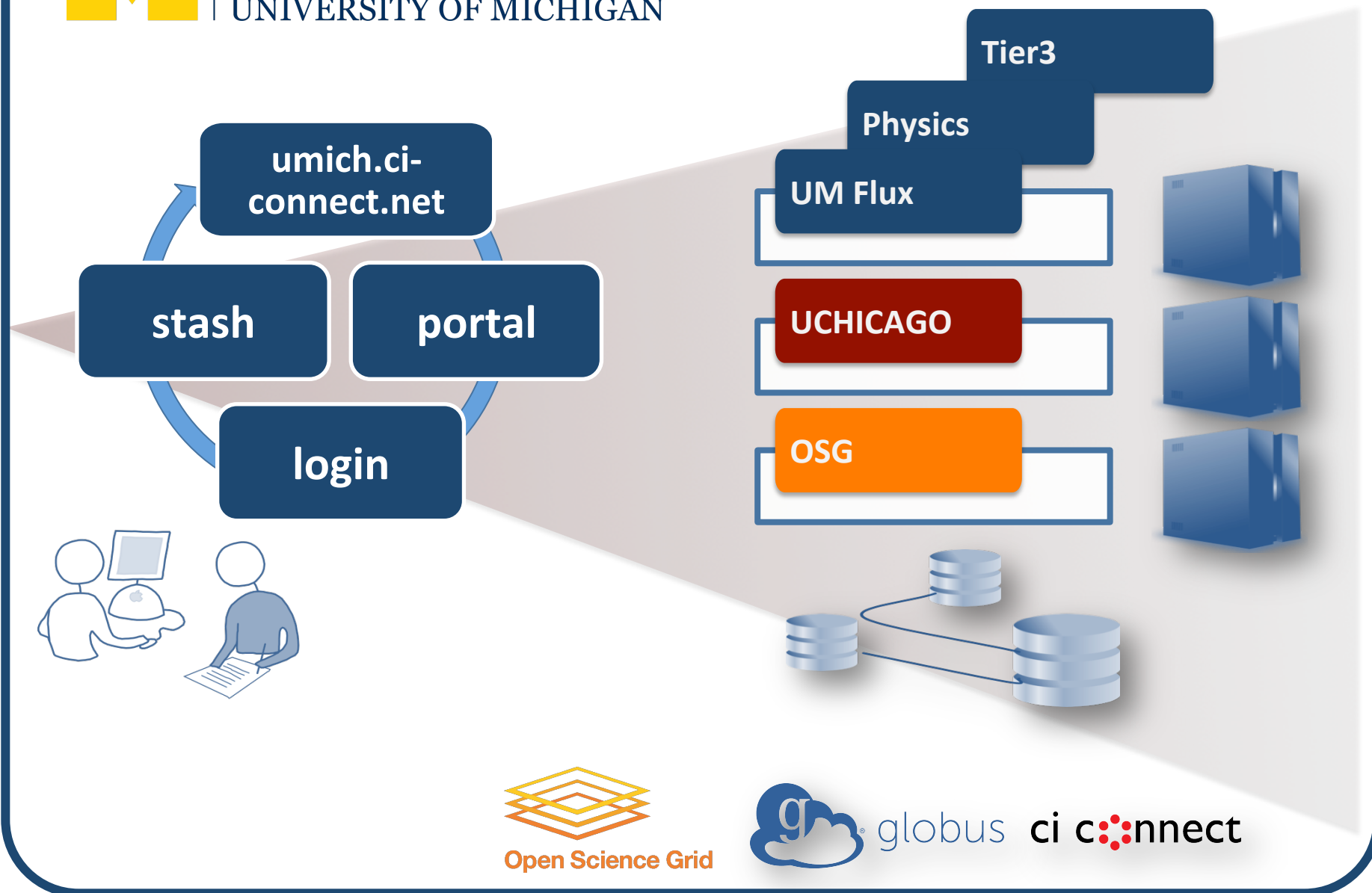


Open Science Grid



globus

ci connect



Cyber Ecosystem Metrics

last year

Campus grids created	3
(Total \geq Level 3 OSG campus grids previously	6)
Campuses bridged	7
Campus grids connected to OSG	2
Off-grid campus research clusters connected	2
XSEDE sites connected	1
OSG Connect users	53
OSG Connect user institutions	12
New OSG Projects	7
Publications forthcoming	5

Challenges

- Encouraging campus cluster users to adopt DHTC
- Emulating home cluster look-n-feel on remote resources, e.g. software utilities
 - /home-cluster, /xsede, /osg environments
- Several workflow options: guiding users to the best fit
- Several data management and access options
- Wrapping off-grid campus clusters with Parrot tools (for software access)
- Compute site data storage or caching

Opportunities, Ideas, Qs

- Develop “standard recipes” for popular workflows, toolkits, science domains
- An OSG http federation and site-level http data caches?
- Giving campus researchers a single environment
 - Multi-user Bosco factory service to reach campus accounts (condo allocations); same for XSEDE
 - Distribute XSEDE campus bridging roll via OASIS
 - OSG user software toolkit distributed via OASIS

Conclusions

- Coming back to...
- The D implies:
 - Ability to connect
 - Ability to bridge
 - Willingness to share
 - Democratization
- Our aim is make it easy, an obvious no-brainer for those with the purse strings



Bonus: a Campus Anthem

Every cycle is sacred

Every cycle is great

If a cycle is wasted....

*God gets quite irate**

